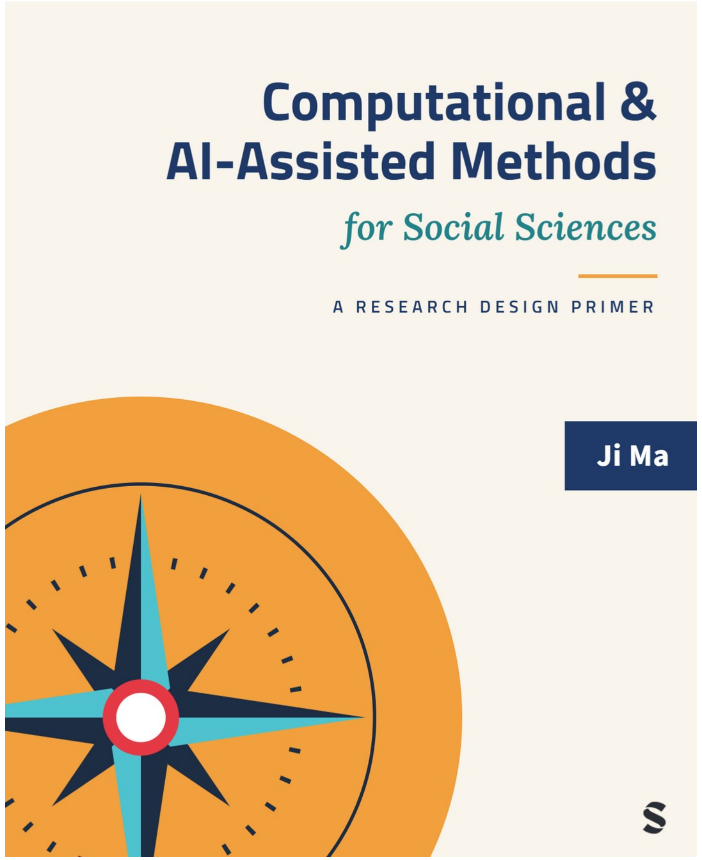
Computational & AI-Assisted Methods for Social Sciences

A Research Design Primer

Chapter 9

Communicating Research



Ji Ma

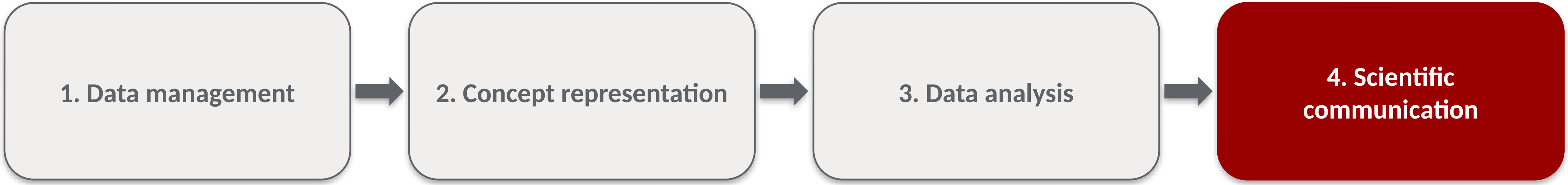
cssprimer.org

Learning objectives

By the end of this chapter, you can:

- Understand scientific communication as a practice spanning the entire research pipeline
- Evaluate transparency practices and apply them to computational projects
- Apply principles of visual perception and design to communicate high-dimensional results honestly
- Recognize CSS- and AI-specific communication challenges, from opacity to reproducibility

Stage 4: scientific communication



The last stage of the pipeline, and the one that makes the other three open to scrutiny.

Takeaway: Communication is where the whole pipeline becomes public and accountable.

§9.1

Scientific communication as a social process

From dissemination to a definition: inspectable, reproducible, interpretable.

Communication was never just the final write-up

- Three classical functions: archival preservation, accessibility, quality control through peer review
- Digital tools collapsed the informal-to-formal publication sequence
- A CSS “finding” includes the data pipeline, measurement, modeling, and interpretation

Takeaway: Every choice that affects the result must be communicated for scrutiny.

Scientific communication makes the entire research pipeline inspectable, reproducible, and interpretable to diverse audiences.

The chapter's working definition

Objectivity is a social achievement, not a personal virtue

- Longino: communities are objective through structures for critical scrutiny
- Inspectability is the precondition for that scrutiny
- Reproducibility is the mechanism for verification
- Interpretability widens the circle beyond specialists

Takeaway: Anything that makes scrutiny harder undermines objectivity, however careful the work.

Open data made the disagreement adjudicable

- Tumasjan et al. (2010): tweet shares tracked German vote shares, 1.65% error
- Jungherr et al. (2012): add back the excluded Pirate Party and error tops nine points
- A meta-analysis: Twitter's predictive power “somewhat exaggerated”

Takeaway: The public API let the community see the pipeline, rerun it, and correct the record.

The post-truth environment raises the stakes

- Misinformation spreads more easily than corrections
- Computational results are easy to misread: dramatic layouts, confident predictions, authoritative AI output
- CSS studies the very platforms that drive misinformation

Takeaway: Rigorous communication is the honest trail that good-faith actors can follow.

§9.2

Making the pipeline inspectable

Declare intentions in advance, share what you actually did, document the AI-assisted steps.

Preregistration separates prediction from postdiction

- Declare hypotheses, methods, and analysis plans before seeing results
- Platforms: OSF for full project management, AsPredicted for a focused template
- Registered reports: peer review before data collection; publication regardless of outcome

Takeaway: Registered reports remove the incentive to publish only positive or surprising results.

Preregistration is not all or nothing

- **Register what you can**: data source, exclusion criteria, outcome, validation strategy
- **Document what you cannot**: the decision *procedure*, not the decision
- **Distinguish** confirmatory from exploratory analyses, and label both

Takeaway: Even a fully exploratory study can lock its pipeline before seeing results.

Open code: the script is part of the argument

- Missing-data handling, exclusions, feature engineering: methodological decisions living in code
- The 2014–2017 baseline: ~1% shared scripts, ~7% shared data, none preregistered
- Shared code compounds: the NTEE classifier was reused for a question it never asked

Takeaway: Sharing thoughtfully beats not sharing: anonymize, aggregate, document restrictions.

Documenting AI-assisted work: three layers, three artifacts

Layer	What it holds	Documentation artifact
Rules	How the tool is used: model ID and version, generation parameters, validation protocol	Project rules file
Knowledge	Stable material the project relies on: full prompts, few-shot examples, output format	Promptbook
Memory	Where the project stands and how it got there: prompt iterations, rationale, validation results	Promptlog and run notes

↗ Chapter 2 templates · cssprimer.org/ch02/

Takeaway: Treat AI-assisted steps with the same documentation rigor as any regression.

§9.3

Making results reproducible

The community must be able to verify, not just see, what you did.

The reproducibility crisis, in three distinct terms

- *Nature* 2016: over 70% failed to reproduce another's experiment; over half failed on their own
- **Reproducibility**: same data, same procedures, same results
- **Replicability**: new data, same hypothesis, same conclusion
- **Repeatability**: the original researcher repeats the analysis

Takeaway: These standards fail for different reasons; diagnose which failure you face.

Freese and Peterson's four forms of replication

	Similar analysis	Different analysis
Original data	**Verifiability**: same data, same analysis, same results?	**Robustness**: do reasonable alternative choices yield the same conclusions?
New data	**Repeatability**: does a similar design on new data repeat the findings?	**Generalizability**: do findings extend to new populations and contexts?

Takeaway: This chapter targets the original-data column: verifiability and robustness.

“Same numbers” is only the minimum standard

- **Outcome** reproducible: identical numerical outputs
- **Analysis** reproducible: minor numeric drift, same interpretation
- **Interpretation** reproducible: different analysis, same conclusion
- Documentation depth runs R1 (description only) to R4 (description + code + data)

Takeaway: The consequential question is whether the conclusion is robust, not just repeated.

Four reproducibility challenges specific to CSS

- Stochastic algorithms: the same code can produce different results without recorded seeds
- API-dependent data: Twitter's 2023 deprecation opened the “post-API age”
- Environment dependencies: package versions and hardware change results
- Prompt sensitivity: outputs shift with wording, model version, temperature

Takeaway: None of these make CSS less scientific; all demand specific technical practices.

Freeze the environment, the seeds, and the data

- Pin dependency versions in lock files: `scikit-learn==1.3.0`, never `>=`
- Isolate each project in a virtual environment
- Containerize when feasible; document hardware when not
- Set, record, and vary random seeds; report sensitivity
- Archive volatile API data at collection, read-only
- Tag the Git commit that produced published results

Takeaway: Together these freeze the execution context so the same inputs give the same results.

Sharing alone does not guarantee reproducing

- Even with materials nominally available, political science reproducibility averages below 50%
- Some journals appoint data editors who verify before publication
- Tiered access preserves verification when full openness is impossible
- Human barrier: self-taught programming without software-engineering habits

Takeaway: Robustness checks belong in the workflow before publication, not after.

AI-assisted analyses: shift the reproducibility boundary

- LLM APIs cannot be pinned: versions deprecate, behavior drifts, temperature zero is not determinism
- Bisbee et al.: after a silent model update, "exactly replicating synthetic data was impossible"
- Archive complete prompt-output pairs; reproduce downstream analyses from the cache
- State in publications where the boundary lies

Takeaway: When the model cannot be frozen, freeze its outputs instead.

§9.4

Making findings interpretable

Design, perception, and honesty for outputs that are high-dimensional, relational, and probabilistic.

CSS outputs are fabricated representations

- Network layouts, embedding projections, and topic distributions have no material pre-existence
- They are *fabricated* through computational and design choices, not *inscribed* like a photograph
- The artifact problem: mistaking products of the visualization for features of the data

Takeaway: Every layout, projection, and color scale is an interpretive decision, not a neutral recording.

The design mindset: decisions before plotting code

- Kirk's hidden thinking: formulate the brief, work with the data, establish editorial thinking
- Then the design solution: encode, lay out, iterate with the audience
- Three guiding principles: trustworthy, accessible, elegant

Takeaway: CSS researchers default to *making*; spend more time thinking before plotting.

The visual encoding hierarchy is measured, not aesthetic

- Accuracy ranking: position, then length, angle, area, color intensity
- Viewers misread circle areas as diameters, shifting ratios by an order of magnitude
- Practical rule: give your most important variable position; reserve color for secondary distinctions

Takeaway: Topic prevalence over time reads better as lines than as stacked areas.

Honest and accessible by design

- Truncated y-axes nearly double the contrast viewers perceive
- Remedy: document the choices, show the same data under alternatives
- Accessibility: colorblind-safe palettes, grayscale-legible figures, alt text
- AI-drafted captions: an accessibility gain only if human-verified

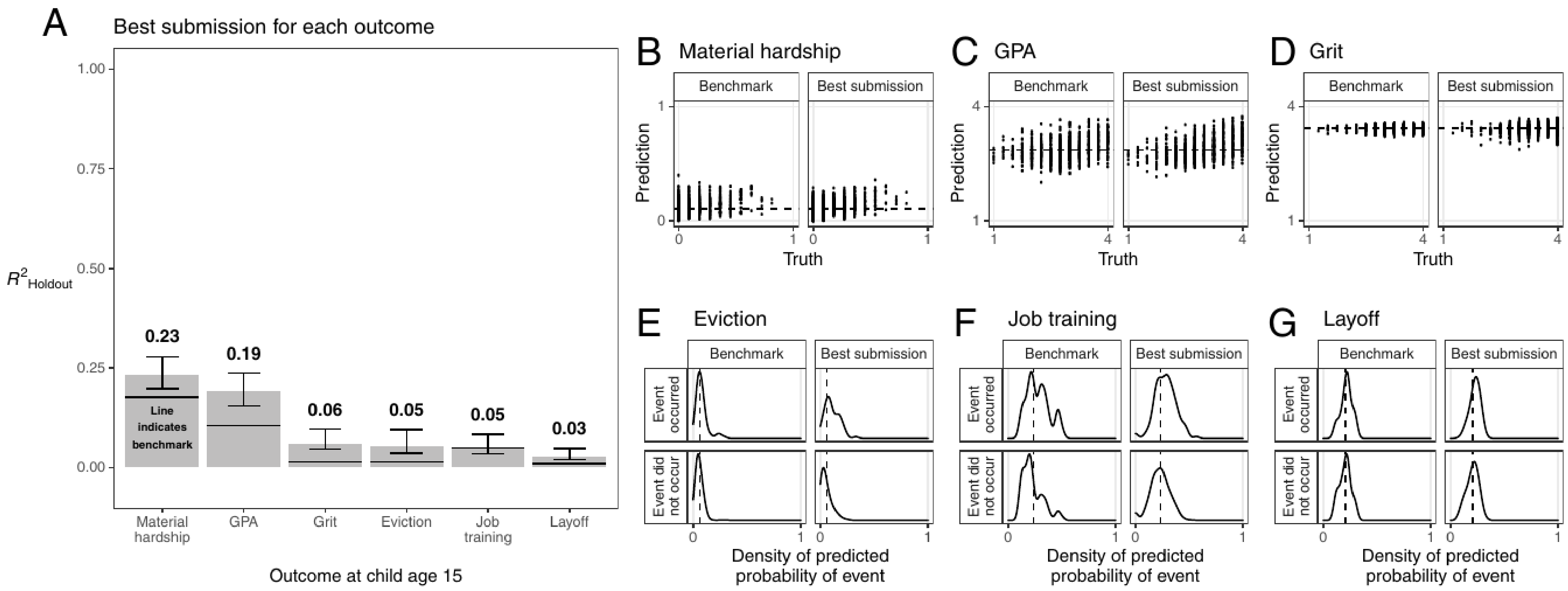
Takeaway: A figure that mystifies part of your audience has failed the people who needed it most.

Three kinds of uncertainty, six visual strategies

- **Statistical** (sampling), **model** (specification), **algorithmic** (seed-to-seed instability)
- Strategies: error bars, icon arrays, hypothetical outcome plots, ensembles, sensitivity tables, opacity
- LLM confidence is not calibrated probability; report empirical accuracy against human coding

Takeaway: A point estimate without its uncertainty makes a precision claim the data may not support.

The Fragile Families Challenge made imprecision the finding

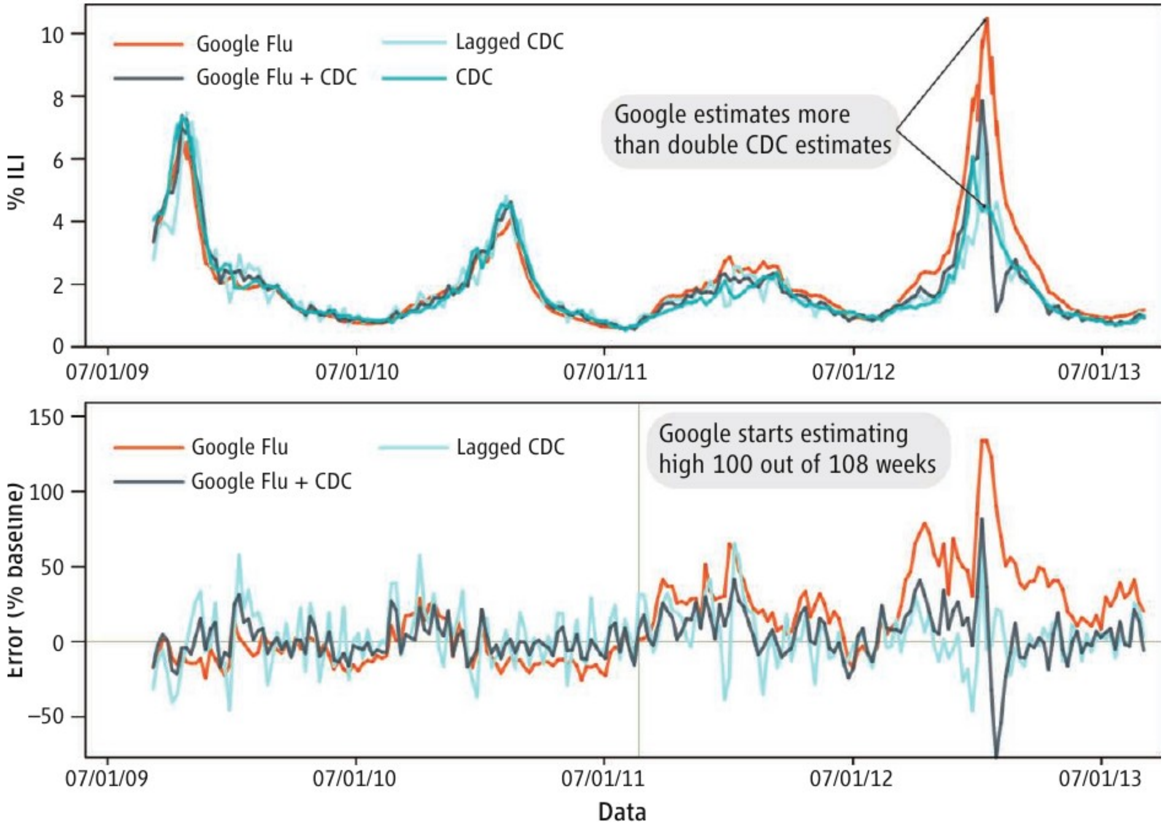


160 teams, six life outcomes: the best hold-out R^2 reaches only ~0.23, with 95% intervals shown.

Source: Salganik et al. (2020), Fig. 3, Proceedings of the National Academy of Sciences.

Takeaway: Showing the confidence intervals, not hiding them, is the figure's central act of honesty.

Google Flu Trends: uncertainty that never traveled with the numbers



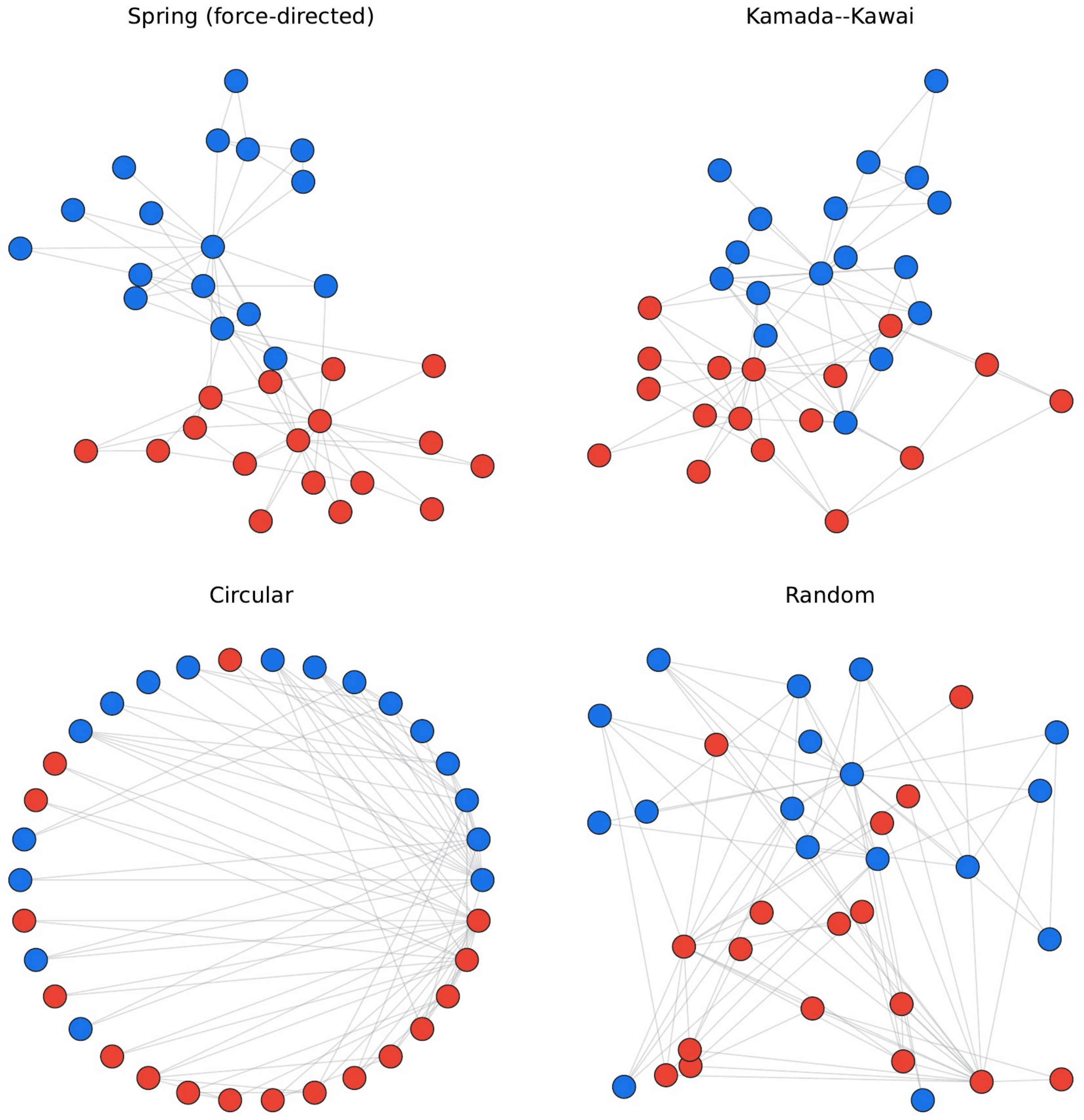
GFT overestimated flu prevalence 100 of 108 weeks after August 2011, doubling CDC's 2013 peak estimate.

Source: Lazer et al. (2014), figure p. 1204, Science. © AAAS. Included for instructional use.

Takeaway: Undocumented algorithms and search terms meant no one could anticipate or audit the drift.

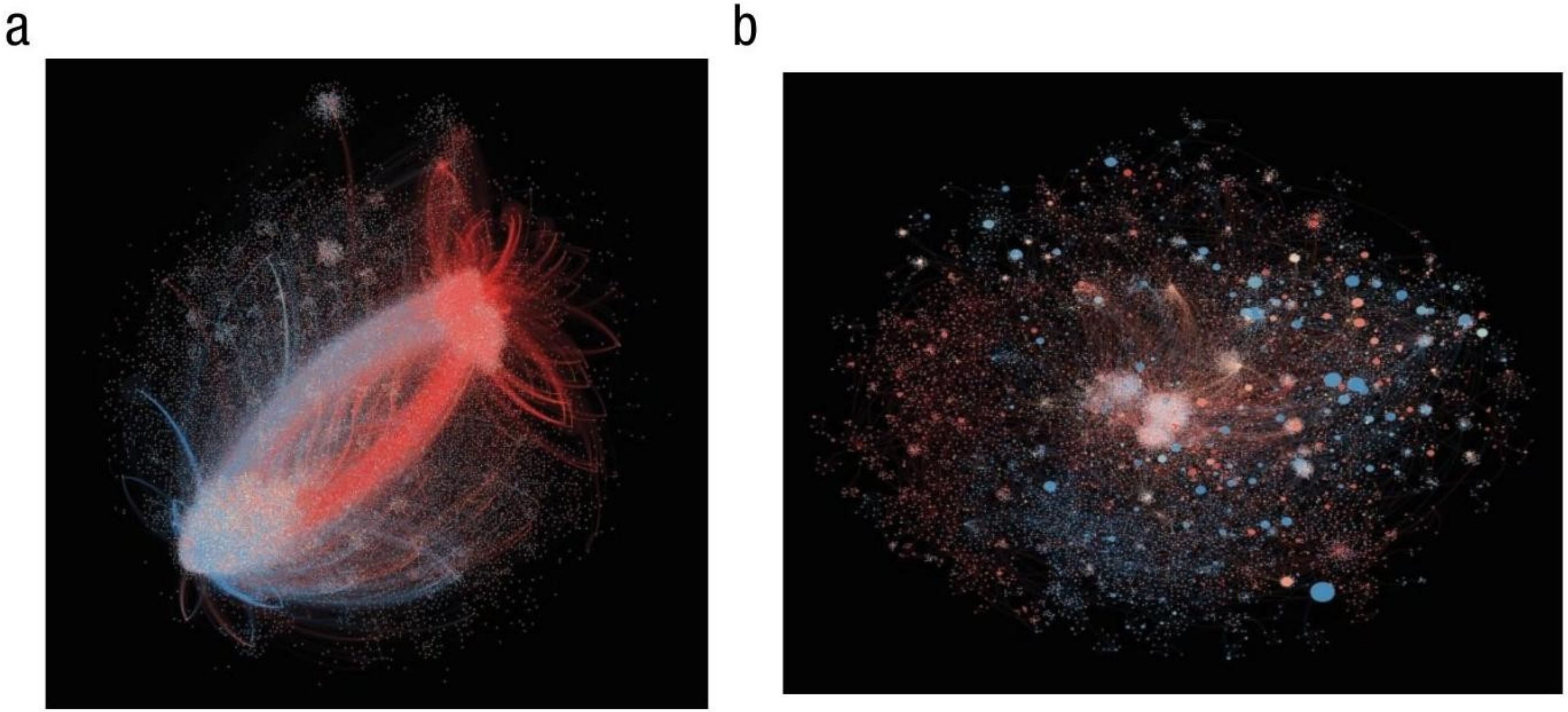
Four layouts, one graph: layout is not data

Zachary's karate club under spring, Kamada–Kawai, circular, and random layouts; only the algorithm changes.



Takeaway: A reader shown only the spring panel would overestimate the community structure.

The same layout dependence in a published finding

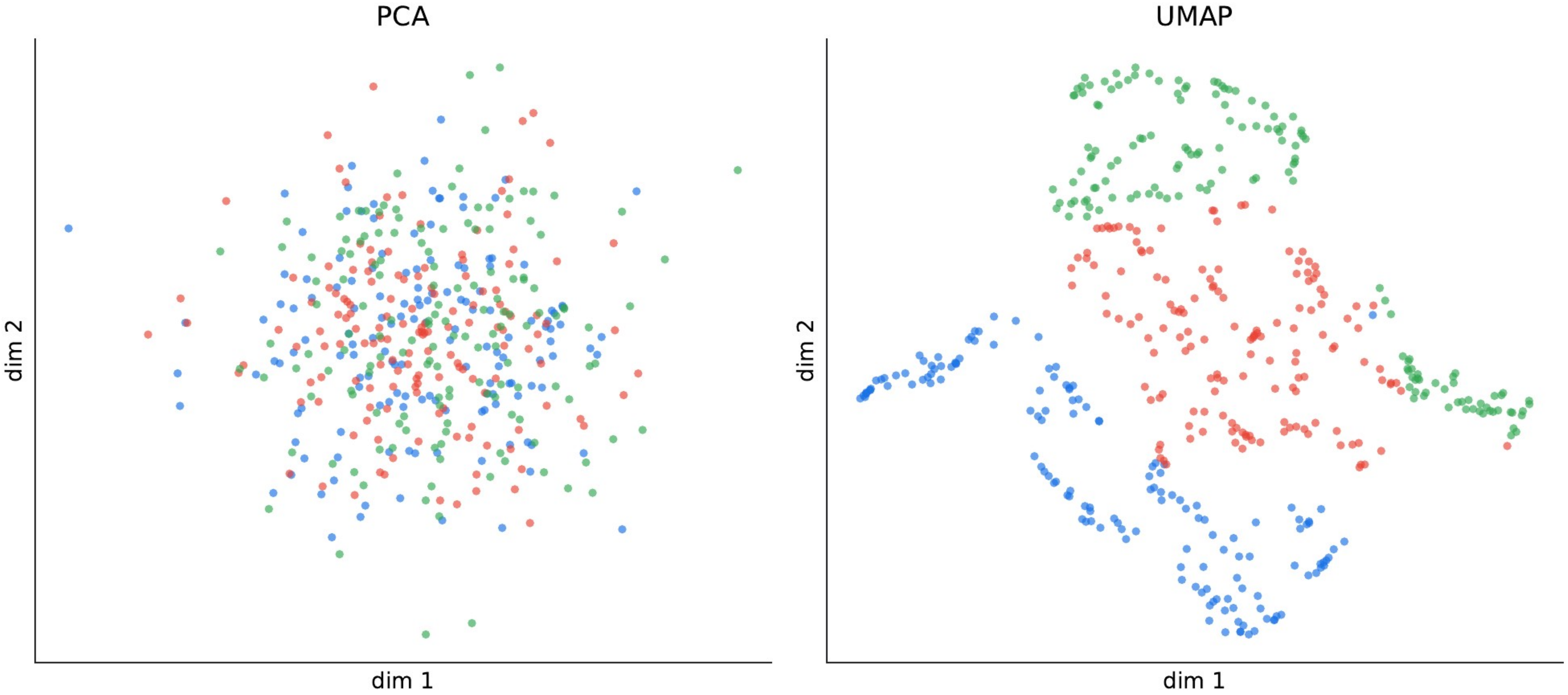


Retweet networks: the 2012 election splits into ideological lobes; the 2014 Super Bowl does not.

Source: Barberá et al. (2015), Fig. 3, Psychological Science.

Takeaway: Pairing a political with a nonpolitical event is good editorial thinking; the unnamed layout is the honesty gap.

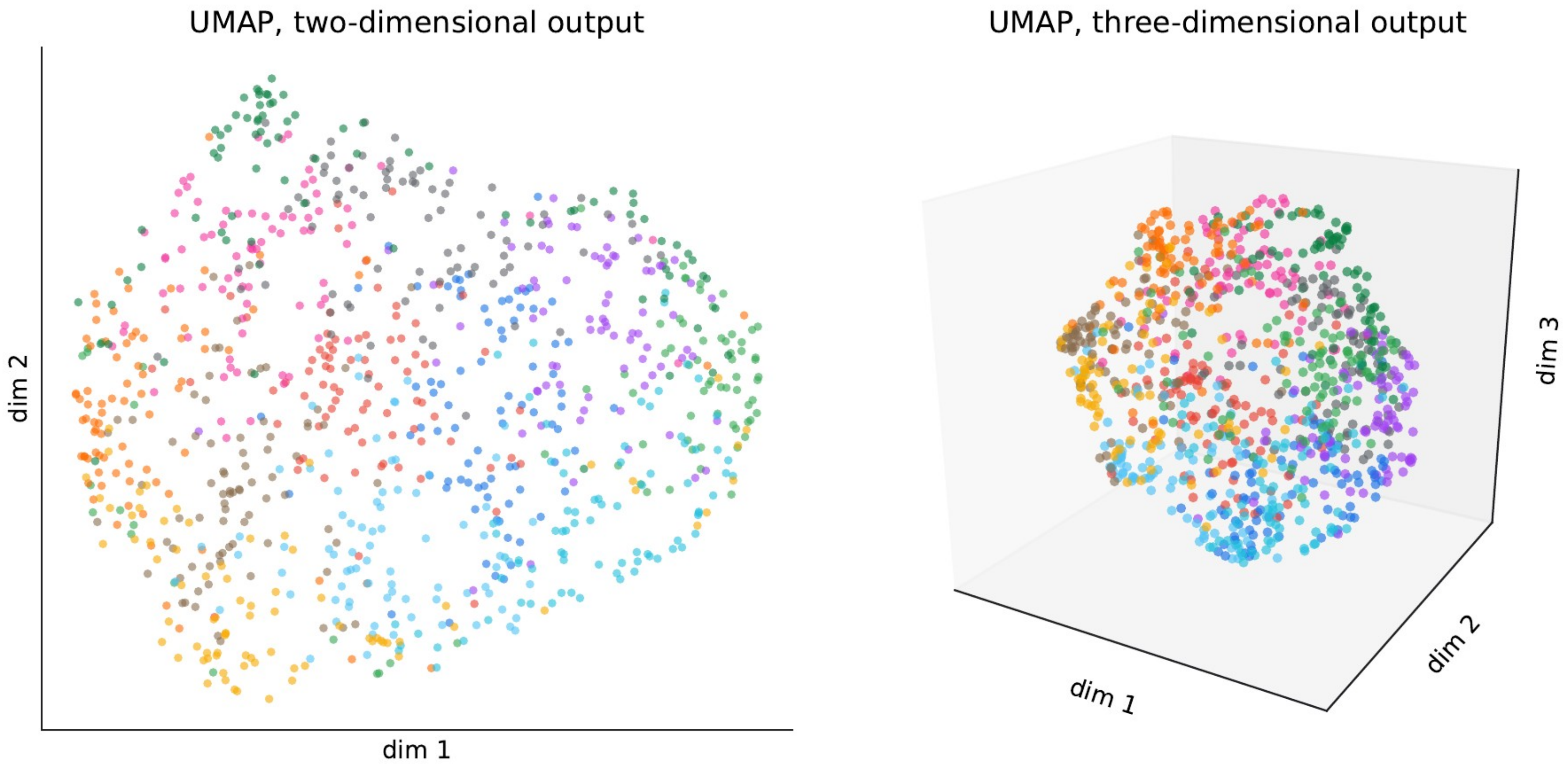
Same data, two projections, opposite answers



A 64-dimensional dataset whose class signal lives on one low-variance axis: PCA shows one blob, UMAP three groups.

Takeaway: The projection algorithm decides what structure your audience gets to see.

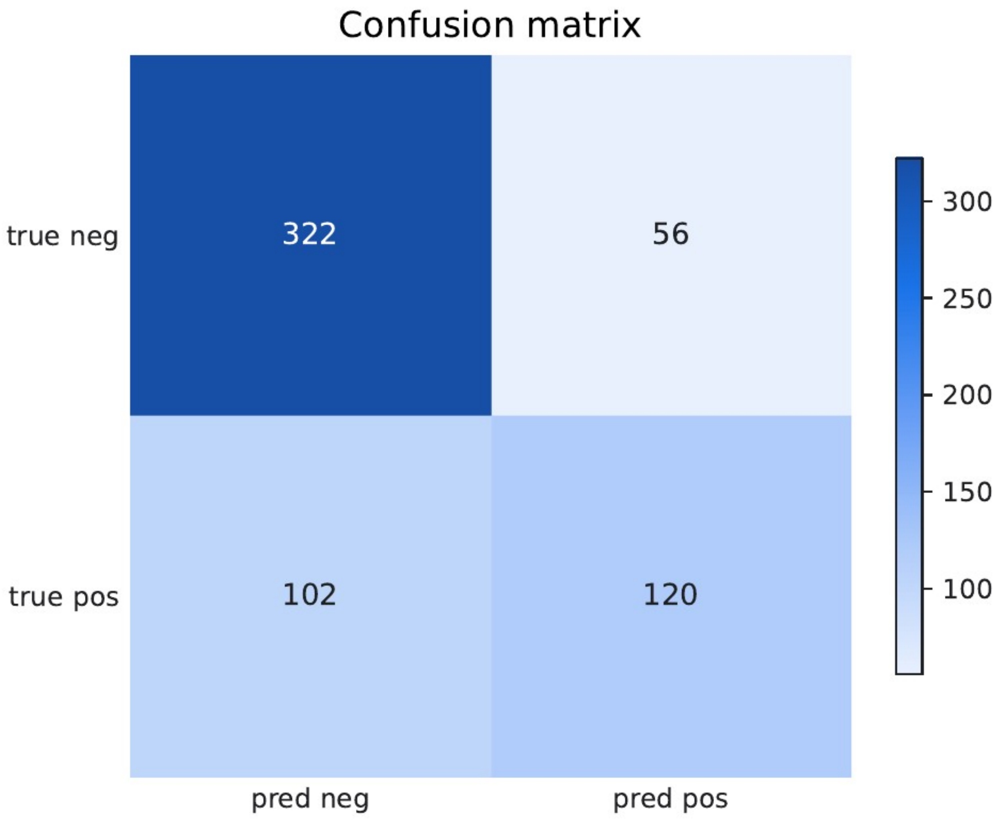
Same data, same algorithm, different output dimensionality

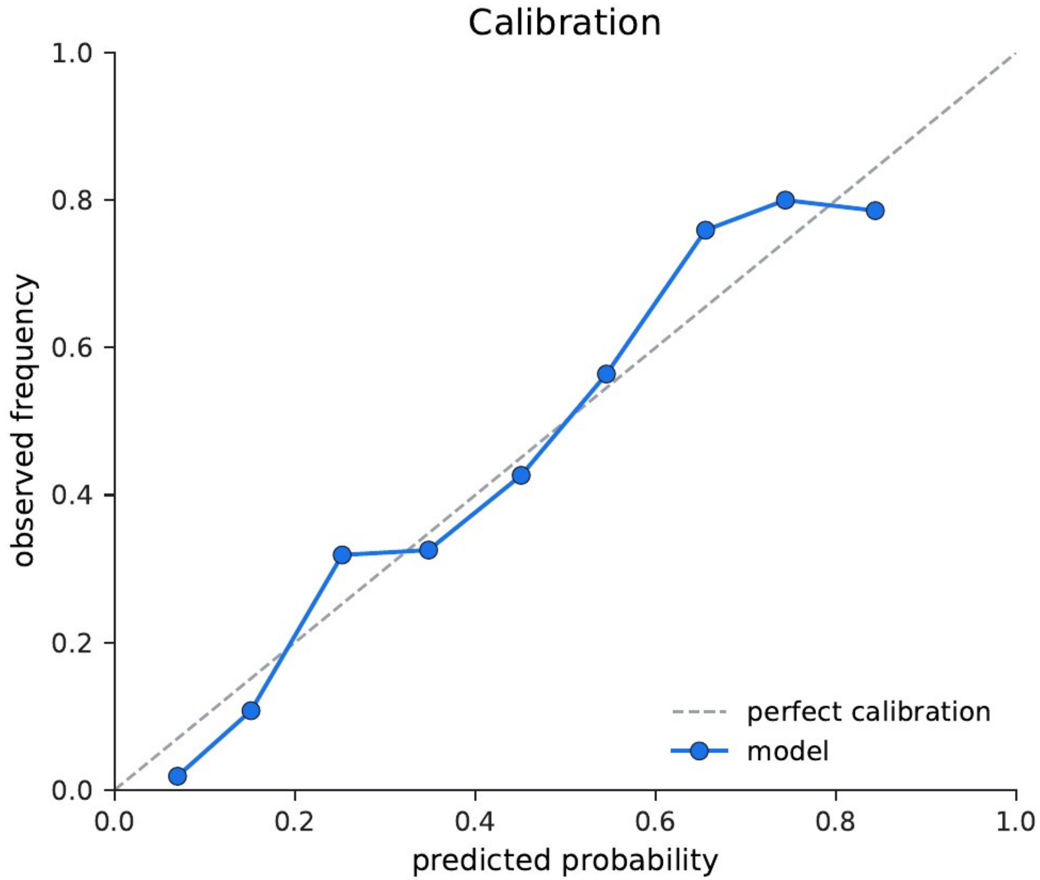


Twelve clusters on a sphere: 3D UMAP recovers the shell; 2D compresses it into an unreadable disk.

Takeaway: You cannot flatten a sphere onto a plane without distortion; neither can UMAP.

Diagnostics for the methods-oriented reader

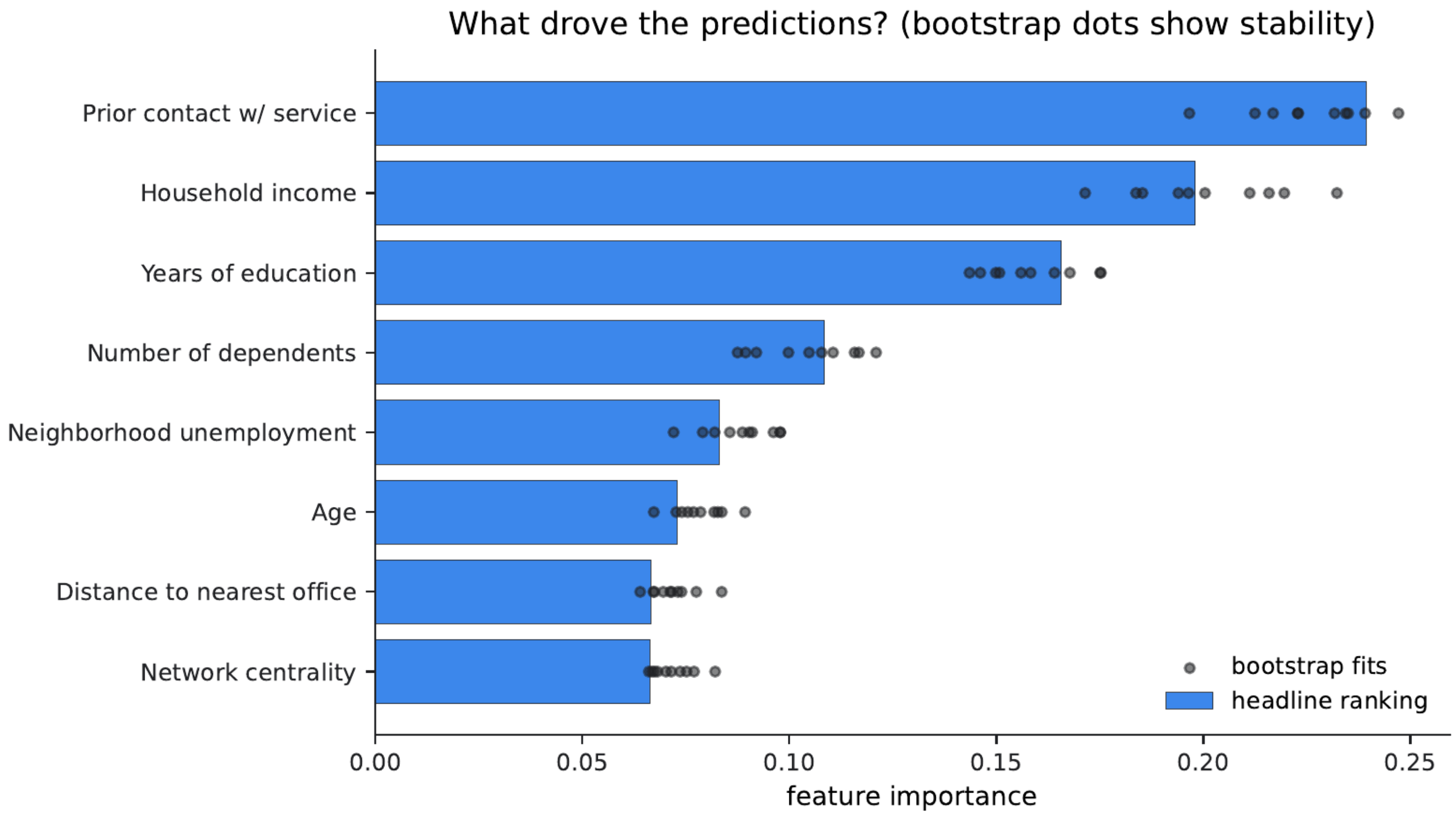




Confusion matrix: 56 false positives, 102 false negatives. Calibration: at a stated 0.65, the actual rate is 0.76.

Takeaway: A single accuracy number would hide both the error asymmetry and the calibration drift.

The same model, translated for the policymaker



Feature importance as ranked bars; dots from ten bootstrap refits show how much each value moves.

Takeaway: The three tiers are stable across refits; the ranks within tiers are not.

Dashboards are living systems, not static figures

- Interactivity lets non-specialists ask their own questions of the data
- Honesty: document the aggregation, filters, and exclusions behind the surface
- Accessibility: static fallbacks, screen readers, keyboard navigation
- Longevity: a 2024 Shiny app may not run in 2030

↗ **Communication tools** · cssprimer.org/tools/communication

Takeaway: Prioritize principles over tools; grammar of graphics outlasts any platform.

Discussion

- Could another researcher see every step of one of your earlier analyses? What is hidden?
- Where does your work fall on the R1–R4 scale, and what is the smallest change that moves it up a rung?
- If your research is exploratory, what can you lock down in advance without being performative?
- Which forms of uncertainty are present in your own work, and which visual strategy fits your audience?

Key concepts from Chapter 9

Scientific communication: Making the entire pipeline inspectable, reproducible, and interpretable to diverse audiences

Preregistration: Declaring hypotheses, methods, and plans before seeing results

Registered report: Peer review before data collection; publication regardless of outcome

Reproducibility: Same data and procedures yield the same results for another researcher

Computational reproducibility: Pinned versions, recorded seeds, archived data, frozen environments

Open data / open code: Sharing data and scripts, with privacy protections, so others can verify and build

AI workflow documentation: Promptbook (Knowledge), promptlog (Memory), model and validation discipline (Rules)

Design mindset: Kirk's four-stage process guided by trustworthy, accessible, elegant

Visual encoding: Position beats length beats angle beats area beats color intensity

Uncertainty communication: Conveying precision and stability: icon arrays, ensembles, sensitivity tables

Materials on the companion site

Studies filtered to scientific communication cssprimer.org/ch09/

The Practice assignment's pre-filtered slice of the studies browser

Communication tools cssprimer.org/tools/communication

Preregistration, environment freezing, visualization, and dashboard tools

CSS Empirical Studies Database cssprimer.org/studies/

The full database; the communication lens has the fewest annotations, which is itself discussable

Where to go next

- Read: Chapter 10, where each property becomes a hands-on exercise on your own project
- Practice: audit one published study with the inspectable–reproducible–interpretable framework
- Browse: the CSS Empirical Studies Database, filtered to scientific communication